Let $P = \{p_1, p_2, \ldots, p_n\}$

$Q = \{q_1, q_2, \ldots, q_n\}$ be 2 sets of corresponding points in $\mathbb{R}^d$.

To find : rotation matrix R and translation vector t such that

$$F(R,t) = \underset{\substack{R \in SO(d) \\ t \in \mathbb{R}^d}}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \| q_i - (Rp_i + t) \|^2$$

where w_i are weights for each point pair

To find t :

$$\frac{dF}{dt} = \sum_{i=1}^{n} 2w_i (Rp_i + t - q_i) = 0$$

$$\therefore 2t \left(\sum_{i=1}^{n} w_i \right) + 2R \left(\sum_{i=1}^{n} w_i p_i \right) - 2 \sum_{i=1}^{n} w_i q_i = 0$$

Taking $\bar{p} = \dfrac{\sum_{i=1}^{n} w_i p_i}{\sum_{i=1}^{n} w_i}$, $\bar{q} = \dfrac{\sum_{i=1}^{n} w_i q_i}{\sum_{i=1}^{n} w_i}$, we get

$$t - \bar{q} + R\bar{p} = 0$$

$$\Rightarrow \boxed{t = \bar{q} - R\bar{p}}$$

Putting t into F,

$$F(R,t) = \sum_{i=1}^{n} w_i \| q_i - Rp_i - \bar{q} + R\bar{p} \|^2$$

$$= \sum_{i=1}^{n} w_i \| R(p_i - \bar{p}) - (q_i - \bar{q}) \|^2$$

Restate problem with zero translation :

$x_i = p_i - \bar{p}$, $y_i = q_i - \bar{q}$

Hence,

$$R = \underset{R \in SO(d)}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \| Rx_i - y_i \|^2$$

Now, $\| Rx_i - y_i \|^2 = (Rx_i - y_i)^T (Rx_i - y_i)$

$= (x_i^T R^T - y_i^T)(Rx_i - y_i)$

$= x_i^T x_i - y_i^T R x_i - x_i^T R^T y_i + y_i^T y_i$

(since $R^T R = I$)

$x_i^T R^T y_i$ is a scalar : $1 \times d \quad d \times d \quad d \times 1 = 1 \times 1$
and hence is equal to its transpose.

$\therefore \; x_i^T R^T y_i = (x_i^T R^T y_i)^T = y_i^T R x_i$

$\therefore \; \| Rx_i - y_i \|^2 = x_i^T x_i - 2 y_i^T R x_i + y_i^T y_i$

$\therefore \; \underset{R \in SO(d)}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \| Rx_i - y_i \|^2$

$= \underset{R \in SO(d)}{\operatorname{argmin}} \left(\sum_{i=1}^{n} w_i x_i^T x_i - 2 \sum_{i=1}^{n} w_i y_i^T R x_i + \sum_{i=1}^{n} w_i y_i^T y_i \right)$

these are not dependent on R and do not matter while minimizing.

$= \underset{R \in SO(d)}{\operatorname{argmin}} \; -2 \sum_{i=1}^{n} w_i y_i^T R x_i$

$= \underset{R \in SO(d)}{\operatorname{argmax}} \sum_{i=1}^{n} w_i y_i^T R x_i$

removing the scalar multiplication.

Now, vectorising $\sum_{i=1}^{n} w_i y_i^T R x_i$, we get

$\sum_{i=1}^{n} w_i y_i^T R x_i = tr(W Y^T R X)$, where :

$$W = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix}_{n \times n} \qquad Y = \begin{bmatrix} - y_1^T - \\ - y_2^T - \\ \vdots \\ - y_n^T - \end{bmatrix}_{d \times n}$$

$$= \text{diag}(w_1, \ldots, w_n)$$

$$X = \begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_n \\ | & | & & | \end{bmatrix}_{d \times n}$$

$\therefore$ We need rotation R which maximizes $tr(WY^TRX)$.

$$tr(WY^TRX) = tr((WY^T)(RX)) = tr(RXWY^T) \qquad [tr(AB) = tr(BA)]$$

Let $d \times d$ covariance matrix $S = XWY^T$. Taking SVD,

$$S = U \Sigma V^T$$

$\therefore$ $tr(RXWY^T) = tr(RS) = tr(RU\Sigma V^T) = tr(\Sigma V^T R U)$

Since V, R, U are orthogonal $\Rightarrow$ $M = V^TRU$ is orthogonal.

$\therefore$ For each column M_j of M, $M_j^T M_j = 1$.

Hence, all numbers M_{ij} are of magnitude $<= 1$.

$$M_j^T M_j = 1 \Rightarrow \sum_{i=1}^{d} M_{ij}^2 = 1 \Rightarrow M_{ij}^2 \leq 1 \Rightarrow |M_{ij}| \leq 1$$

$$\therefore tr(\Sigma M) = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_d \end{bmatrix} M = \sum_{i=1}^{d} \sigma_i M_{ii} \leq \sum_{i=1}^{d} \sigma_i$$

(where Σ is a diagonal matrix and $\sigma_1, \sigma_2, \ldots, \sigma_d \geq 0$.)

$\therefore$ To maximize $tr(\Sigma M)$, $M_{ii} = 1$.
Since M is orthogonal $\Rightarrow$ to maximize $tr(\Sigma M)$, $M = I$.

$\therefore M = V^T R U = I \Rightarrow V = RU$
$\Rightarrow \boxed{R = VU^T}$

Hence proved mathematically that Procrustes alignment gives the best aligning transform between point clouds with known correspondances.